Online Appendix

Appendix II: Identifying African American Inventors in Patent Data

As I mentioned, it is very difficult to identify the race of a patentee because patent records did not record it, with only one exception since 1790. The first systematic attempt to identify African American patentees was an effort by the Patent Office, which undertook surveys in 1900 and 1913. The objective of the surveys was to locate African American patentees whose achievements would be featured in the 1900 Paris World's Fair and to commemorate scientific achievements by African Americans in the 50 years following the end of the Civil War. Directed by one of the lead examiners, Henry E. Baker, surveys were sent to 9,000 of the approximately 12,000 patent attorneys and agents. Responses to the survey were collected and analyzed by Baker and published in several formats (Baker, 1913, 1917, 1921). A subset of the original responses were donated to Carter G. Woodson, a noted historian, and, in turn, donated by him to the Library of Congress. The Baker data extend from 1834 to 1917. The investigation in this paper required that the data be extended to 1940.

A first strategy to extend the data set was to include patents obtained in 1913 and beyond by inventors already in the data set. These data were collected using the European Patent Office (EPO) search engine, which is searchable by name from 1920. Google Patent Search, which can also conduct historical searches, became available after 2004, when these data were originally collected. Google Patent Search misses some historical patents, and the EPO search is more reliable.

One strategy for identifying additional black inventors would be to match patentees from USPTO data to Census data. This method should work for inventors who lived and patented in the same place. However, this procedure fails, because African Americans during this period did not patent where most African Americans live, as Table 2 shows. Before 1940, most African American inventors obtained patents in northern states rather than in southern states. Unlike today, specific addresses were not reported by the Patent Office, just the city or town in which the inventor resided or from which he or she applied for a patent. It is difficult to find a unique first- and last-name match using census data, because of the proximity of first and last names of African American inventors to those of other inventors, especially inventors of British origin. Eight patentees were identified as African American using this method. Only with significant additional biographical data does this method work, and these data are available for few inventors in the data set. And if additional biographical data were introduced, the selection problem would be of greater concern, because biographical information is available for only the most famous and prolific inventors.

Another strategy would be to match common names given to African Americans to patent data. A three-pronged strategy in the spirit of Fryer and Levitt (2004) and Bertrand and Mullainathan (2004) was executed but was unsuccessful in identifying black patentees. This mechanism is described below. A second-best method would be to match known black inventors to names in the patent data. This method was significantly more successful in producing matches. The second method and its limitations are described in the text.

An index of black names for the period 1870 to 1940 was constructed from census data in two ways. The first strategy answered the question: conditional on being black, which names are most likely to be observed? Random samples of black ("Negro"), "mulatto", and "colored" heads of households from the 1870, 1900, and 1920 censuses were drawn for the District of Columbia, Georgia, Michigan, and New York. From these samples, frequencies were calculated for first and

last names separately. There were 14 first names and 11 last names that appeared more frequently than the median frequency and were included in the index.

The second strategy answered the question: conditional on observing a certain name, what is the likelihood that the person is black? First and last names of blacks ("Negroes") and whites were extracted from the 5% IPUMS sample of the 1870 census. Unlike the above samples, names were not restricted to heads of households. From these samples, frequencies were calculated for first and last names separately and by race for names occurring at least 80 times. Among blacks, there were 27 first names and 20 last names that appeared more frequently than the median frequency for whites or were a larger share of the total names than the black share of the total population and were included in the index.

A third approach was an extension of the second approach and answered the question: conditional on having a name widely adopted by African Americans following the end of slavery, what is the probability that the person is black? This strategy was intended to take advantage of a well-known practice among African Americans of adopting the first and last names of presidents, e.g., George Washington, or famous people in the black community, e.g., Booker T. Washington, as first and middle names. The entire 1900 census was used and also was not restricted to heads of households.

These approaches yielded largely similar results from which I constructed an index of "black names." Results were nearly identical with respect to surnames.

Yet the composite index was unable to predict matches in the 1880 census sample of the 690 individuals identifying their occupation as "inventor." I was able to predict a small number of black inventors but, with the exception of George Washington Carver and George Washington Murray, not ones that could be matched to a patent. The index significantly underpredicted matches to black inventors and overpredicted matches to white inventors in New England, particularly those born in England, as was the case with the first census-based approach. Additional location and biographical data would have been required to obtain unique first- and last-name matches. In general, these methods are more suitable for the current rather than historical period.

This highlights a problem associated with occupation identification and reporting among inventors. Many identify themselves as machinists or artisans or engineers rather than inventors, irrespective of race. Thomas Edison, among other "great inventors" who are alive and active as inventors, does not appear in the 1880 sample.

The final strategy to extend the Baker data set was to construct a broad-based data set of African American inventors, i.e., potential patentees, and to match the resulting data to patent data. Among the historical and contemporary sources used to create a pool of potential patentees were searches of historical newspapers, including obituaries, e.g., from the Ohio Historical Society Newspaper online database and newspaperarchive.com; correspondence from Carter G. Woodson, Henry E. Baker, and patent survey participants (Library of Congress); the Garrett Morgan Papers; historical and contemporary directories of African American medical doctors, scientists, and engineers, e.g., academic journals, including the *Journal of Economic History* and the *Journal of Negro History*; historical and contemporary biographies of African American inventors and general biographies, e.g., *Great Negroes Past and Present*; and programs of exhibitors in the African American sections or exhibitions of historical fairs, including the "Exhibit of American Negroes" at the 1900 Paris World's Fair, the 1904 "Great Negro Fair" in Raleigh, North Carolina, and the 1933 Chicago World's Fair "Negro Day." Newspaper and obituary searches and programs of exhibitions allowed the identification of lesser known inventors. A complete list of sources appears in a companion paper. Not all inventors and others in the pool of potential patentees were matched to patent records and were dropped from the data set. Others were dropped if there was not a unique first- and last-name match, e.g., James Young in the patent data. Ultimately, while second best, this

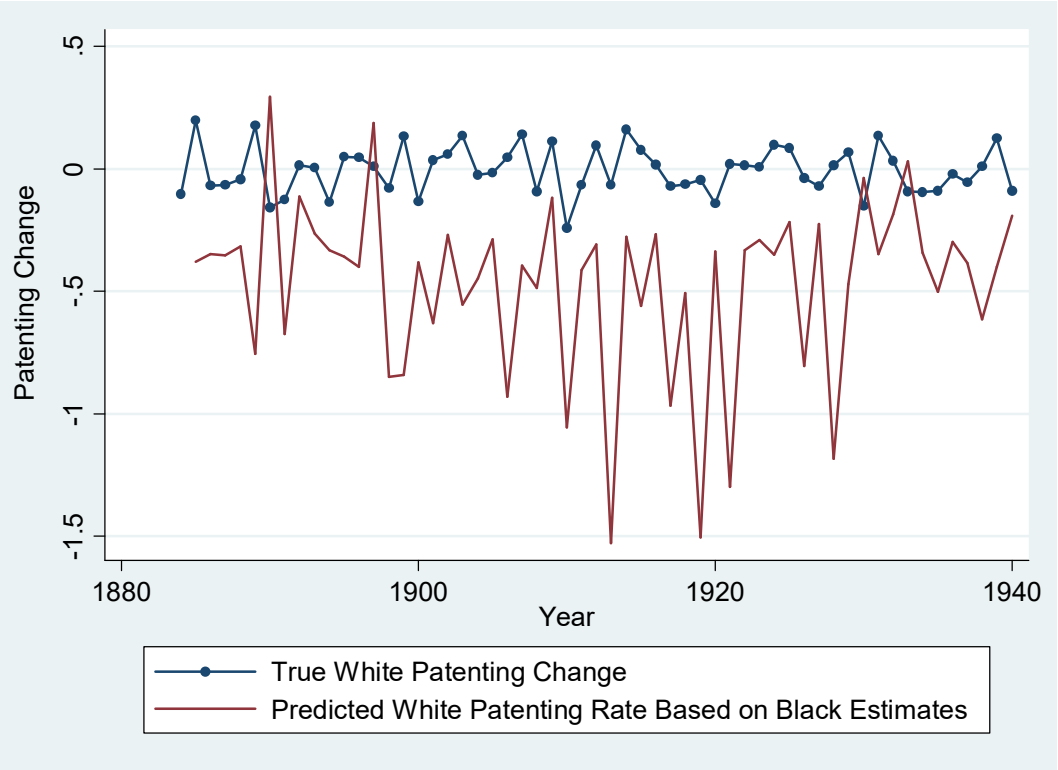
process provides a more systematic and less ad hoc means of recovering black patentees to extend the data set. This summary is adapted from Cook (2011).

Appendix III. Estimating the Effect on Productivity Using Black Parameters

Let us consider a second counterfactual exercise. How much lower would inventive activity in the U.S. have been if all inventors operated under violence-related conditions? We can address this question by taking parameter estimates from estimation of equation 3 in the black subsample and using them in the white subsample. As can be seen in Figure 4, patent output over the period 1882 to 1940 would have been nearly 1% per year, or 40% over the period, lower and would have displayed significantly more volatility for white inventors, who constitute the overwhelming majority of inventors at that time. Like Abadie and Gardezeabal (2003), I find that volatility seems to increase in the presence of greater violence. Of course, the comparison using black parameters with the white sample should be interpreted with caution, because it not only reflects the evolution of violence but also pre-violence differences in determinants related to patent or economic activity. There is an imperfect mapping between technological progress and patent activity, and there are other factors that would change in the white subsample, e.g., illiteracy rates. Nonetheless, the results are suggestive that the rate of technical change in the U.S. may have been substantially lower in the absence of the rule of law affecting both races. Further, concomitant improvements in living standards may have increased much more slowly.

Appendix III.

Figure 4. Predicted White Patent Activity Using African American Estimates



Source: Cook (2004), Author's calculations.
Note: Figure 4 is estimated from Equation (1).